

Zhiyong Lu, Ph.D.

CONTACT INFORMATION	DHHS/NIH/NLM/NCBI 8600 Rockville Pike Building 38A, Room 1003A Bethesda, MD 20894	<i>Tel:</i> (301) 594-7089 <i>Fax:</i> (301) 480-2288 <i>E-mail:</i> zhiyong.lu@nih.gov http://www.ncbi.nlm.nih.gov/CBBresearch/Lu
RESEARCH INTERESTS	Bioinformatics, Biomedical Informatics Text Mining (Named Entity Recognition; Biomedical Information Retrieval) Machine Learning, Data Mining	
EMPLOYMENT	Associate Investigator, National Center for Biotechnology Information (NCBI), 2009 - Staff Scientist, National Center for Biotechnology Information (NCBI), 2007 - 2009	
EDUCATION	Ph.D. in Bioinformatics, University of Colorado Health Sciences Center, 2007 PhD Thesis: Text Mining on GeneRIFs, advised by Lawrence Hunter M.S. in Computer Science, University of Alberta, 2003 MS Thesis: Predicting Protein Sub-cellular Localization from Homologs Using Machine Learning Algorithms, supervised by Duane Szafron and Russell Greiner B.S. in Computer Science, Nanjing University, 2001	
HONORS AND AWARDS	Special Service Award, for outstanding investigation and computational analysis of online pharmaceutical information resulting in a higher quality resource delivered to consumers, National Library of Medicine, 2010 Individual Special Service Award, for excellence in performance of PubMed log analyses and development of new and valuable user services, National Library of Medicine, 2009 AMIA Annual Symposium 2008 Outstanding Paper Submission, 2008 Best performance in automatic protein-protein interaction extraction task, BioCreative II global competition, 2007 Conference (PSB 2007) Travel Award, Int'l Society of Computational Biology, 2007 Conference (ISMB 2006) Travel Award, Int'l Society of Computational Biology, 2006 Conference (PSB 2006) Travel Award, Int'l Society of Computational Biology, 2006 Conference (ISMB 2005) Travel Award, Int'l Society of Computational Biology, 2005 Professional Development Award, University of Colorado Health Sciences Center Student Senete, 2005 Conference (PSB 2005) Travel Award, Int'l Society of Computational Biology, 2005 Conference (Rocky 03) Travel Award, University of Colorado Center of Computational Biology, 2003 Undergraduate (merit-based) Student Scholarship, Nanjing University, 1997 - 2001	

PROFESSIONAL
SERVICES

Conference and Workshop Organizing:

Organizing committee, the International Biomedical Informatics Summit of Peking University, Beijing 2011

Co-organizer for the third BioCreative global challenge

Local co-chair, the 3rd international BioCreative Workshop, USA 2010

Poster co-chair, the 3rd international BioCreative Workshop, USA 2010

Grant study section and reviews:

External reviewer, NSERC discovery grants, 2011

Review and Referee Work for Journals:

Bioinformatics

BMC Bioinformatics

BMC Medical Genomics

Journal of Biomedical Informatics (JBI)

ACM Transactions on Information Systems (TOIS)

IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)

*DATABASE - The Journal of Biological Databases and Curation
Information*

Scientific Program Committee for Conferences:

ACL-HLT BioNLP Shared Task 2011

ACL-HLT BioNLP workshop, Portland, USA 2011

AMIA 2010 Annual Symposium, Washington DC USA, 2010

International Conference for Machine Learning and Applications, Washington, DC, 2010

ACL-HLT BioNLP workshop, Sweden, 2010

AMIA 2009 Annual Symposium, San Francisco USA, 2009

International Joint Conferences on Artificial Intelligence (IJCAI), Pasadena, CA, 2009

ISMB/ECCB 2009 Stockholm, Sweden, 2009

The Asia Pacific Bioinformatics Conference (APBC), 2009

NAACL-HLT BioNLP workshop, Boulder, CO 2009

ACL-HLT BioNLP workshop, Columbus, Ohio 2008

NAACL-HLT BioNLP workshop, Brooklyn, NY 2006

Pacific Symposium on Biocomputing (PSB), Maui, HI 2006

MENTORSHIP

Ph.D. Thesis Committee:

Andrew Fox, Ph.D. in Computational Biology, Tufts University, 2010

Co-supervised PhD Students:

Craig Murray, University of Maryland at College Park, January 2008 - September, 2010

Research Fellows:

Rezarta I. Dogan, Ph.D. August 2008 - present

Aurelie Neveol, Ph.D. October 2008 - present

Jiao Li, Ph.D. March 2010 - present

Others:

Minlie Huang, Ph.D. Visiting Assistant Professor from Department of Computer Science, Tsinghua University, January 2010 to June 2010

**TEACHING
EXPERIENCE**

Conference Tutorial: Introduction to biomedical natural language processing, *The Asia Pacific Bioinformatics Conference (APBC) 2009*, Beijing, China. January, 2009.

05/06-11/06 University of Colorado School of Medicine Graduate Student Supervisor
Supervise summer intern students and rotation graduate students

08/04-11/04 University of Colorado School of Medicine Teaching Assistant
BIOI 7711: Introduction to Bioinformatics (graduate level class)

09/02-12/02 University of Alberta Teaching Assistant
CMPUT 366: Introduction to Artificial Intelligence

01/02-04/02 University of Alberta Teaching Assistant
CMPUT 201: Practical Programming Methodology

09/01-12/01 University of Alberta Teaching Assistant
CMPUT 101: Introduction to Computing

**OTHER SERVICES
AND
MEMBERSHIPS**

Member of the ACM Special Interest Group on Information Retrieval (SIGIR), 2010 -
Member of the Association of Computational Linguistics (ACL), 2008 -
Founding member of the ACL Special Interest Group, Biomedical Natural Language Processing (BIOSIGMED), 2008 -
Member of the International Society of Computational Biology (ISCB), 2005 -
Student Representative in Department Faculty Meeting, University of Colorado, 2004
Treasurer-elect, CS Graduate Student Association, University of Alberta, 2003
Councilor-elect, Graduate Student Association, University of Alberta, 2003
Member of the American Association for Artificial Intelligence (AAAI), 2002

**INVITED
LECTURES**

Mining literature for accelerating knowledge acquisition and discovery, stadman symposium on bioinformatics, National Institutes of Health, December, 2010

Text Mining for Omics, Genomics and Proteomics Interest Group, George Washington University, March 2009.

Evaluating IR Techniques for Enhancing MEDLINE Retrieval, Computational Bioscience Program, University of Colorado School of Medicine, December, 2008.

Evaluating Relevance Ranking Strategies for MEDLINE Retrieval, NLM Informatics Series, December, 2008

Evaluating and Enhancing NCBI Textual Resources, CBCB Seminar, University of Maryland College Park, November 2008

Evaluating and Enhancing NCBI Textual Resources, AI Seminar, Department of Computer Science, University of Alberta, April 2008

Text Mining on GeneRIFs, CLIP Colloquium, University of Maryland College Park, November 2007

Text Mining for Biology, Center for Bioinformatics, Peking University, June 2007

Text Mining for Biology, School of Life Sciences, Suzhou University, February 2007

Extracting Protein Transport Information from GeneRIFs, Bioinformatics supergroup, University of Colorado at Boulder, March 2006

Case Study: Protein Translocation, Biomedical Ontology Session Discussion Panel, Pacific Symposium on Biocomputing, January 2005

PEER REVIEWED
PUBLICATIONS

1. **Lu Z***, PubMed and beyond: a survey of Web tools for searching biomedical literature, *Database*, 2011
2. Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvernin V, Church DM, Dicuccio M, Federhen S, Feolo M, Fingerman IM, Geer LY, Helmberg W, Kapustin Y, Landsman D, Lipman DJ, **Lu Z**, Madden TL, Madej T, Maglott DR, Marchler-Bauer A, Miller V, Mizrachi I, Ostell J, Panchenko A, Phan L, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Shumway M, Sirotkin K, Slotta D, Souvorov A, Starchenko G, Tatusova TA, Wagner L, Wang Y, Wilbur WJ, Yaschenko E, Ye J. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 2011
3. Neveol A, Islamaj-Dogan R, **Lu Z***, Quality, efficiency, satisfaction: a study of semi-automatic semantic annotation of PubMed queries, *J Biomed Inform*, 2010
4. Islamaj-Dogan R, and **Lu Z***, Learning to predict document keywords from a user perspective, *Bioinformatics*, 2010 26(21):2767-75
5. **Lu Z***, and Wilbur WJ, Overview of BioCreative III Gene Normalization, in *Proceedings of the BioCreative III workshop*, Maryland, 2010.
6. Islamaj-Dogan R, Yang Y, Huang M, Neveol A, and **Lu Z***, Identifying protein-protein interactions in biomedical articles, in *Proceedings of the BioCreative III workshop*, Maryland, 2010
7. Neveol A, and **Lu Z**, Automatically integration of drug indications from multiple ehealth records, to appear in *Proceedings of the 1st ACM International Health Informatics Symposium*, 2010.
8. Islamaj-Dogan R, Neveol A, **Lu Z**, A novel textual representation scheme for identifying clinical relationships in patients records, to appear in *Proceedings of the The Ninth*

9. Neveol A, Islamaj-Dogan R, **Lu Z**^{*}, Author keywords in biomedical journal articles, *to appear at AMIA*, 2010. (selected for journal publication).
10. Huang M and **Lu Z**, Learning to annotate scientific publications, in *Proceedings of the 23rd International Conference on Computational Linguistics*, Coling 2010.
11. Mork JG, Bodenreider O, Demner-Fushman Dina, Dogan RI, Lang FM, **Lu Z**, Neveol A, Peters L, Shooshan SE, Aronson AR, Extracting Rx Information from Clinical Narrative, *JAMIA*, 2010
12. Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvernin V, Church DM, Dicuccio M, Federhen S, Feolo M, Geer LY, Helmberg W, Kapustin Y, Landsman D, Lipman DJ, **Lu Z**, Madden TL, Madej T, Maglott DR, Marchler-Bauer A, Miller V, Mizrachi I, Ostell J, Panchenko A, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Shumway M, Sirotkin K, Slotta D, Souvorov A, Starchenko G, Tatusova TA, Wagner L, Wang Y, John Wilbur W, Yaschenko E, Ye J. Database resources of the National Center for Biotechnology Information, *Nucleic Acids Research*, Nov 2009
13. Dogan RI, Murray GC, Neveol A, **Lu Z**^{*}, Understanding PubMed user search behavior through log analysis, *Database*, 2009(0): bap018
14. **Lu Z**^{*}, Xie N, Wilbur WJ: Identifying related journals through log analysis, *Bioinformatics*, 25(22):3038-9, 2009. (currently used in NLM's Journals database)
15. **Lu Z**^{*}, Wilbur WJ, McEntyre JR, Iskhakov A, Szilagyi L: Finding query suggestions for PubMed, in *proceedings of the AMIA Annual Symposium 2009*. (selected for journal publication; currently used in PubMed)
16. Jin F, Huang M, **Lu Z**, Zhu X: Towards Automatic Generation of Gene Summary, in *proceedings of the BioNLP 2009 Workshop*, Boulder, USA 2009
17. Neveol A, Kim W, Wilbur WJ, **Lu Z**^{*}: Exploring two biomedical text genres for disease recognition, in *proceedings of the BioNLP 2009 Workshop*, Boulder, USA 2009
18. Chen S, Huang M, **Lu Z**: Summarizing Documents by Measuring the Importance of a Subset of Vertices within a Graph, in *proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, Italy, 2009
19. **Lu Z**^{*} and Wilbur WJ: Improving accuracy for identifying related PubMed queries by an integrated approach, *Journal of Biomedical Informatics*, 2008.
20. **Lu Z**^{*}, Kim W, Wilbur WJ: Evaluation of Query Expansion Using MeSH in PubMed, *Information Retrieval*, Volume 12, Issue 1, pages 69-80 2009.

^{*}Corresponding author

21. **Lu Z***, Kim W, Wilbur WJ: Evaluating Relevance Ranking Strategies for MEDLINE Retrieval, *JAMIA*, 2009. (*recognized as one of the 9 outstanding papers from a total of 457 papers submitted to the AMIA 2008 Annual Symposium, and invited for publication in JAMIA*).
22. Morgan AA, **Lu Z**, Wang X, Cohen AM, Fluck J, Ruch P, Divoli A, Fundel K, Leaman R, Hakenberg J, Sun C, Liu H, Torres R, Krauthammer M, Lau WW, Liu H, Hsu C, Schuemie M, Cohen KB, Hirschman L: Overview of BioCreative II Gene Normalization, *Genome Biology*, 2008, 9(Suppl 2):S2
23. Baumgartner WA, Jr, **Lu Z**, Johnson HL, Caporaso JG, Paquette J, Lindemann A, White EK, Medvedeva O, Cohen KB, Hunter L: Concept recognition for extracting protein interaction relations from biomedical text, *Genome Biology*, 2008, 9(Suppl 2):S8
24. Bethard S, **Lu Z***, Martin JH, Hunter L: Semantic Role Labeling for Protein Transport Predicates, *BMC Bioinformatics*, 9:277, 2008
25. Hunter L, **Lu Z**, Firby J, Baumgartner WA, Jr., Johnson HL, Ogren PV, Cohen KB: OpenDMAP: An open-source ontology-driven concept analysis engine, with applications to capturing knowledge regarding protein transport, protein interaction, and cell-specific gene expression, *BMC Bioinformatics*, 9(1):78, 2008
26. Baumgartner W, **Lu Z**, Johnson HL, Caporaso JG, Paquette J, Lindemann A, White EK, Medvedeva O, Cohen KB, and Hunter L: An integrated approach to concept recognition in biomedical text, In *Proceedings of the Second BioCreative Challenge Evaluation Workshop*, 2007
27. **Lu Z***, Cohen KB, and Hunter L: GeneRIF Quality Assurance as Summary Revision, in *Proceedings of the Pacific Symposium on Biocomputing* 12:269-280, 2007
28. Caporaso JG, Baumgartner W, Kim H, **Lu Z**, Johnson H, Medvedeva O, Lindemann A, Fox L, White EK, Cohen KB, Hunter L: Concept Recognition, Information Retrieval, and Machine Learning in Genomics Question-Answering, in *Proceedings of The Fifteenth Text Rtrieval Conference*, (TREC 2006)
29. **Lu Z**, Bada M, Ogren P, Cohen KB, and Hunter L: Improving Biomedical Corpus Annotation Guidelines, in *Proceedings of the Joint BioLink and 9th Bio-Ontologies Meeting*, 5 August 2006; Fortaleza, Brazil
30. **Lu Z***, Cohen KB, and Hunter L: Finding GeneRIFs via Gene Ontology Terms, in *Proceedings of the Pacific Symposium on Biocomputing*, 3-7 January 2006, Maui, Hawaii. 2006:52-63
31. Johnson HL, Cohen KB, Baumgartner WA, **Lu Z**, Bada M, Kester T, Kim H, and Hunter L: Evaluation of Lexical Methods for Detecting Relationships Between Concepts From Multiple Ontologies, in *Proceedings of the Pacific Symposium on Biocomputing*, 3-7 January 2006, Maui, Hawaii. 2006:28-49

32. **Lu, Z*** and Hunter, L: GO Molecular Function Terms are Predictive of Subcellular Localization, in *Proceedings of the Pacific Symposium on Biocomputing*, 4-8 January 2005, Big Island of Hawaii. 2005:151-161
33. Szafron D, Lu P, Greiner R, Wishart DS, Poulin B, Eisner R, **Lu Z**, Anvik J, Macdonell C, Fyshe A, Meeuwis D: Proteome Analyst: custom predictions with explanations in a web-based tool for high-throughput proteome annotations. *Nucleic Acids Research* 2004 Jul 1, 32:W365-71
34. **Lu Z**, Szafron D, Greiner R, Lu P, Wishart DS, Poulin B, Anvik J, Macdonell C, Eisner R: Predicting subcellular localization of proteins using machine-learned classifiers. *Bioinformatics*. 2004 Mar 1; 20(4): 547-56
35. Szafron D, Greiner R, Lu P, Wishart DS, **Lu Z**, Poulin B, Eisner R, Anvik J, Macdonell C: Proteome Analyst Transparent High-throughput Protein Function Prediction. In *Proceedings of the International Conference on Machine Learning, Workshop on Machine Learning in Bioinformatics*, August 2003, Washington DC, USA
36. Sander J, Qin X, **Lu Z**, Niu N and Kovarsky A: Automatic Extraction of Clusters from Hierarchical Clustering Representations, in *Proceedings of The 7th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, April 2003, Seoul Korea.

REFEREED
ABSTRACTS

37. Islamaj Dogan R, Neveol A, **Lu Z**, A textual representation scheme for identifying clinical relationships in patient records, *The Ninth International Conference on Machine Learning and Applications*, Washington D.C., USA 2010
38. Islamaj-Dogan R, Murray GC, Neveol A, **Lu Z**, Characterizing User Search Behavior in PubMed, *AMIA 2010 Annual Symposium*, Washington D.C., USA 2010.
39. Li j, **Lu Z**, Zhu X, Chen JY, A Computational Method for Mining the Relationships between Diseases, Proteins and Drugs from Molecular Network and PubMed, *AMIA 2010 Annual Symposium*, Washington D.C., USA 2010.
40. Neveol A, Kim W, **Lu Z**, Scenario-specific information retrieval in the biomedical domain, *AMIA 2010 Annual Symposium*, Washington D.C., USA 2010.
41. Liu J, Huang M, Zhu X, Wilbur WJ, and **Lu Z**, A novel learning based system for assisting gene indexing, *BioCreative III workshop*, Bethesda, Maryland, USA, 2010.
42. Islamaj-Dogan R, Murray GC, Neveol A, **Lu Z**, Towards efficient search tools for biomedical databases: Characterizing user search habits and recognizing their information needs, *The Eighteenth International Conference on Intelligent Systems for Molecular Biology (ISMB 2010)*, Boston, USA, 2010.
43. Li j, **Lu Z**, Zhu X, Chen JY, Systematic evaluation of disease gene identification in molecular interaction networks, *The Eighteenth International Conference on Intelligent Systems for*

Molecular Biology (ISMB 2010), Boston, USA, 2010.

44. Mork JG, Bodenreider O, Demner-Fushman D, Dogan RI, Lang FM, **Lu Z**, Neveol A, Peters L, Shooshan SE, Aronson AR, NLM's I2B2 Tool System Description, *the Third i2b2 Workshop on Natural Language Processing Challenges for Clinical Records*, November, San Francisco, 2009.
45. Murray GC, Lin J, Wilbur WJ, **Lu Z**, Users' Adjustments to Unsuccessful Queries in Biomedical Search, *the 9th Joint Conference on Digital Library*, Austin, Texas, 2009.
46. **Lu Z**, Kim W, Wilbur WJ, Evaluating Relevance Ranking and Query Expansion for MEDLINE Retrieval, *the 6th Annual Rocky Mountain Bioinformatics Conference*, Colorado, 2008.
47. **Lu Z**, Baumgartner WA, Caporaso JG, Cohen KB, Hunter L. Finding informative sentences in full-text journal articles, *the 4th Annual Rocky Mountain Bioinformatics Conference*, Colorado, 2006.
48. Caporaso JG, Baumgartner WA, Kim H, **Lu Z**, Johnson HL, Medvedeva O, Lindemann A, Fox L, White EK, Cohen KB, Hunter L. Information retrieval, question-answering, machine learning, and concept recognition in TREC Genomics 2006, *the 4th Annual Rocky Mountain Bioinformatics Conference*, Colorado, 2006.
49. **Lu Z**, Baumgartner W, Bada M, Firby J, Ogren P, Cohen K, Hunter L. Frame-based Analysis of Complex Relationships in Biomedical Text: Annotating and Extracting Protein Transport Information in GeneRIFs, *the 14th International Conference on Intelligent Systems for Molecular Biology*, Fortaleza, Brazil, 2006
50. **Lu Z**, Hunter L. Mining Protein Transport Data from GeneRIFs, *the 3rd Annual Rocky Mountain Bioinformatics Conference*, Snowmass, Colorado, 2005
51. **Lu Z**, Ogren P, Dolbey A, Hunter L. PTKB: Protein Transport Knowledge Base, *the 13th International Conference on Intelligent Systems for Molecular Biology*, Detroit, Michigan, 2005
52. **Lu Z**, Hunter L. GO Molecular Function Terms are Predicative of Subcellular Localization, *the 2nd Annual Rocky Mountain Bioinformatics Conference*, Snowmass, Colorado, 2005
53. **Lu Z**, Using Proteome Analyst for Protein Subcellular Localization Predication, *the 1st Annual Rocky Mountain Bioinformatics Conference*, Snowmass, Colorado, 2004
54. **Lu Z**, Szafron D, Lu P, Wishart DS, Poulin B, Anvik J, Macdonell C, Eisner R. Using Proteome Analyst for Protein Subcellular Localization Predication, *the 3rd International Conference of the Canadian Proteomics Initiative*, Vancouver, BC, 2003
55. **Lu Z**, Wu X, Greiner R. Protein Sequence Classification in Proteome Analyst, *the 10th International Conference on Intelligent System for Molecular Biology*, Edmonton, Alberta,

2002

PAPERS IN
PREPARATION

Huang M, Neveol A, **Lu Z**^{*}, Recommending MeSH Terms for Annotating Biomedical Articles, *submitted*

Murray GC, Kim W, **Lu Z**, Wilbur WJ, Machine Learning for Query Term Removal: Assisting users with failed searches in PubMed, *submitted*